NEXT-GENERATION PROTEIN SEQUENCING ADVANCES VARIANT IDENTIFICATION

Proteins are the essential workers of biology. The 20,000-plus proteins and over 1 million proteoforms that exist in human cells have a massive purview and agenda, solving critical problems to maintain wellness by accomplishing important tasks that enable almost every biological system to run smoothly and successfully.[1] There is much we have uncovered about these biomolecules, and yet with every answer come many more questions.

Researchers are tackling the challenge of fully understanding the proteome. The inherent complexity of proteins is one part of this challenge. Another is prohibitive technologies: traditional methods of scrutinizing proteins, such as mass spectrometry, are expensive, require a large amount of lab space, and tend to be accessible only to those with certain levels of institutional investment. But thanks to a new high-performance platform that leverages next-generation protein sequencing™, we may be on the verge of a deeper understanding of these precious cellular components.

Enter Platinum,™ the Next-Generation Protein Sequencer™ produced by Quantum-Si to solve a critical challenge around sensitive and unambiguous amino acid detection. A first-of-its-kind benchtop sequencer capable of single molecule resolution, Platinum unites biology, chemistry, and semiconductor technology to help life scientists see, learn from, and harness knowledge about proteins in ways that can potentially transform basic biology research and medical applications.

The future of next-generation sequencing for proteins has the potential to be as revolutionary as next-generation sequencing for nucleic acids. In the field of genomics, next-generation sequencing technology enables cost-effective parallel sequencing of millions to billions of DNA fragments. It has found applications in investigating molecular-level variation in cancer, rare genetic diseases, and complex microbiomes.[2]

Molecular-level information about individual variation is also fundamental to understanding proteins' role in health and disease. Functional proteins can take chemically different forms, depending on mutations during transcription, splices during translation, or modifications or cleavage after translation, among other sources of variation. Because detecting these changes is difficult, the diversity of proteoforms in the human proteome remains largely unmapped.[1, 3]

The impact of proteoform variation on health and disease is a work in progress. Next-generation sequencing provided by Platinum provides resolution at the amino acid level, which is necessary to identify the complex landscape of variation. With this new tool, Quantum-Si continues its decade-long passion to democratize protein sequencing and make the technology available to all scientists.[4]

THE PLATINUM DIFFERENCE

Conventional methods to examine protein variation involve trade-offs between ease of use and level of detail in protein identification data (figure 1). Researchers use immunoassays, such as Western blot and ELISA (enzyme-linked immunoassays), to quickly identify proteins in a sample. But these techniques require customized reagents in the form of high-quality antibodies, and they need additional analysis to get protein sequence or individual amino acid information.

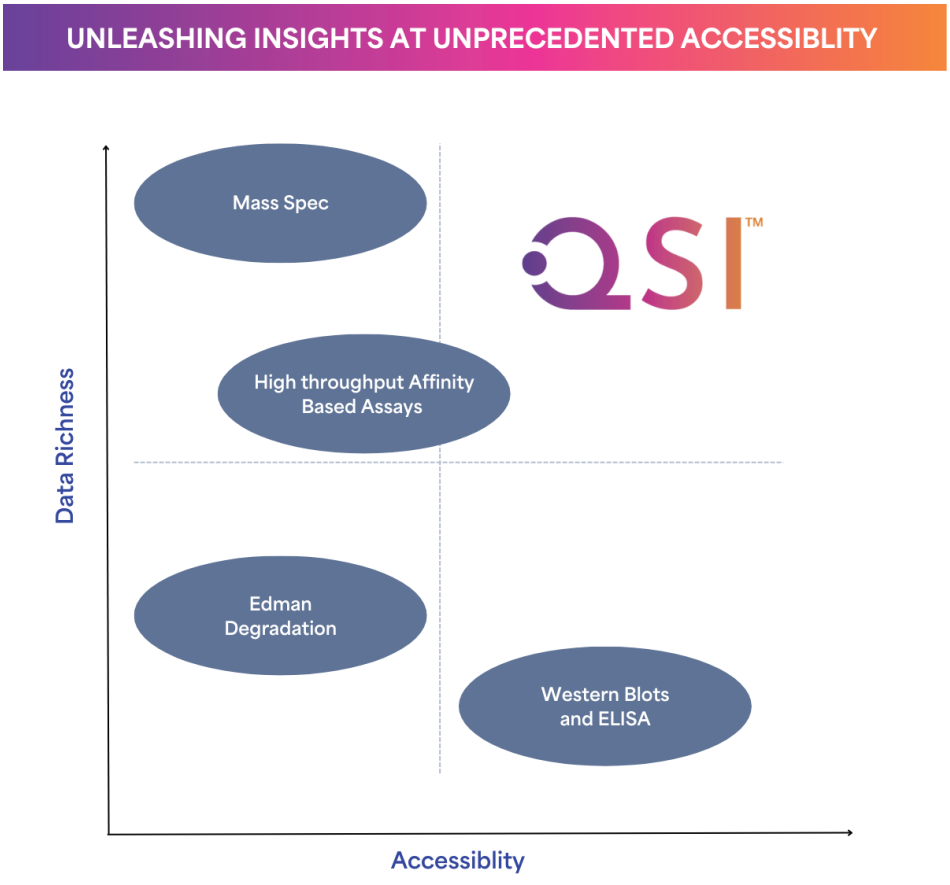

Figure 1: Protein sequencing and identification technologies can involve a trade-off between ease of use and richness of the resulting data.

Courtesy of Quantum-Si

Mass spectrometry (MS) is an established technique for protein identification at varying levels of detail. Researchers use this technique to analyze intact proteoforms or their small peptides.[5] The naturally low abundance of proteoforms means that signal detection can be challenging, However, MS also provides sequence information by inference: it reports a molecular mass to charge (m/z) ratio, which can obscure detail when variants have the same m/z ratio.

Researchers using MS for protein analysis often send their samples to core facilities. This inconvenience is amplified for many scientists who do not have access to expensive, large infrastructure or the needed expertise, which includes complex bioinformatics approaches to analyze their data.

Platinum, on the other hand, is an easy-to-install benchtop sequencer with automated data analysis delivering results that are easily interpreted. It identifies a protein sequence through direct recognition of amino acids. This single molecule, amino acid level of resolution enables researchers to explore proteoforms and modifications, uncovering what is actually there rather than what could be there. This information is foundational to understanding how protein variation impacts biological function.

Most crucially, this relatively low-cost, next-generation protein sequencer is intended to seamlessly complement lab workflows and methodologies that scientists use every day. Neil L. Kelleher, a molecular bioscientist and director of the Chemistry of Life Processes Institute at Northwestern University, is excited about how Platinum serves as an effective counterpart to other systems he uses in his workflow. He is collaborating with Danielle Tullman-Ercek, professor of chemical and biological engineering at Northwestern, who engineers self-assembling proteins for various therapeutic, industrial, and environmental applications. The work requires controlling which proteoforms are secreted at what time and from what location in engineered cells.

Dr. Tullman-Ercek's lab at Northwestern is using Platinum as a test bed to characterize the dynamics of protein production and combine those sequences with a data set from top-down mass spectrometry, Kelleher says. "Most people view mass spectrometry–based proteomics and these new protein sequencing methods with single-molecule resolution as competing," he adds. "I like the word cooperation. The future demands that we provide more clarity about human proteins for the improved detection of disease and for the understanding of fundamental mechanisms that evolved through the billion or so years of evolution on the planet."

Platinum's sequencing power resides in using *N*-terminal amino acid (NAA) recognizers to sequence peptides from a target protein. Enzymes in solution, called aminopeptidases, sequentially remove individual NAAs to expose subsequent amino acids for recognition. The recognizers are also labeled with a fluorescent dye, and their on-and-off binding to NAAs creates a pulsing pattern with characteristic fluorescence and kinetic properties that can be used to derive sequence information (figure 2).[6]

Quantum-Si researchers sequenced three different proteins with Platinum and sent the same proteins to a core facility for sequencing via MS. Both techniques identified each protein, but Platinum provided a simpler workflow to identify specific peptide fragments in each protein.[7]

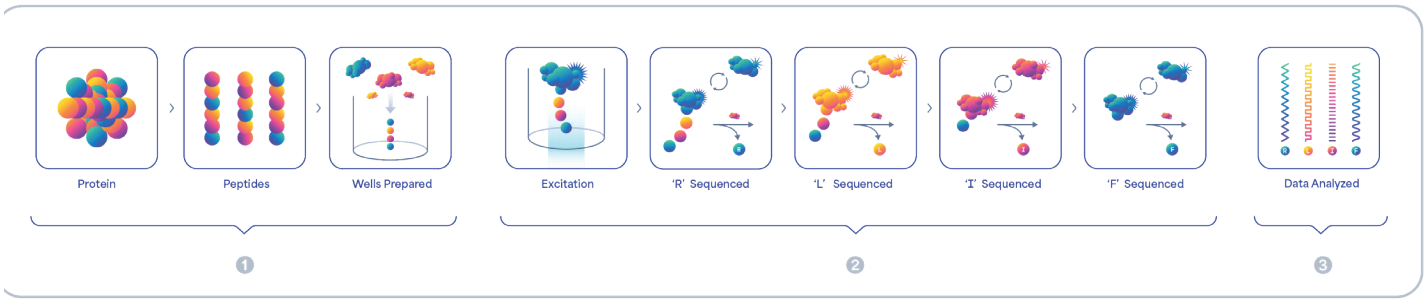

Figure 2: The workflow for the Platinum protein sequencer involves three phases: Prepare, Sequence, and Analyze. During Prepare (1), a target protein is digested into peptides and the peptide library is loaded into a proprietary semiconductor chip along with a collection of recognizers specific to N-terminal amino acids. During Sequence (2), excitation of the recognizers, combined with the recognizers' on-and-off binding, creates a pulsing pattern with characteristic fluorescence and kinetic properties that can be used during Analysis (3) to derive sequence information.

Courtesy of Quantum-Si

A UNIVERSE OF APPLICATIONS

Platinum provides protein sequencing at single-molecule resolution in the first-of-it's-kind benchtop instrument.

Courtesy of Quantum-Si

Applications explored with Platinum include distinguishing proteins in a mixture and identifying variants.

When detecting individual proteins in the complexity of a biological sample, researchers typically use antibodies specific to their target protein. Immunoassays cannot provide sequence information without additional analysis, however. Quantum-Si researchers tested Platinum's ability to identify individual proteins in a five-component mixture. Peptide identification for the mixture was identical to that of each protein sequenced individually, which demonstrates that Platinum can identify multiple proteins in a single-sample preparation and sequencing run.[8]

To demonstrate how Platinum can be employed in researching diseases that are impacted by changes in protein sequence, Quantum-Si scientists distinguished three variants of the virus that causes COVID-19 according to differences in the amino acid sequence of viral spike proteins. The ability to rapidly identify viral variants enables scientists to understand their potential impact on public health

and develop effective treatments and vaccines.[9]

With Platinum, all researchers can have a next-generation protein sequencer with single-molecule detection in their lab. This means proteomics is accessible to nonexperts, including genomics researchers who want to gather protein information too. Specialists in genomics can now easily expand their work into multiomics studies. They no longer have to assume that protein production follows genetic instructions like a textbook: researchers can now use genetic sequences and protein sequences to see the actual gene-protein relationship in a cell.

FUTURE OF PROTEOMICS RESEARCH

Platinum is poised to tackle proteomics' great challenge of mapping proteoforms[10], akin to the Human Genome Project in studying all the DNA in the human body. The presence of a post-translational modification (PTM) on an amino acid will alter the binding kinetics of the recognizers and result in a unique kinetic binding signature. To illustrate this on Platinum, Quantum-Si researchers interrogated modifications of the amino acid arginine that are of a type linked to cancer and cardiovascular conditions. They chose two modifications that are difficult to detect using MS because the modifications have different chemical structures with the same mass. With Platinum, understanding single-molecule amino acid variations and modifications is now possible.[11]

"I'm fired up," Kelleher says. "You have a major opportunity to transform the area of proteomics, which includes scaling proteoform analysis."

The question is whether tools can be built that match the "challenge of our biology with the scale of the solutions," he adds. "I hope that we can get a Human Genome Project-scale effort across the three sectors of government, industry, and academia to really fundamentally transform our knowledge of the human proteome."

REFERENCES

1. Ruedi Aebersold et al., "How Many Human Proteoforms Are There?," *Nat. Chem. Biol.* 14, no. 3 (March 1, 2018): 206–14, https://doi.org/10.1038/nchembio.2576.
2. Heena Satam et al., "Next-Generation Sequencing Technology: Current Trends and Advancements," *Biology* 12, no. 7 (July 13, 2023): 997, https://doi.org/10.3390/biology12070997.
3. Lloyd M. Smith and Neil L. Kelleher, "Proteoforms as the Next Proteomics Currency," *Science* 359, no. 6380 (March 9, 2018): 1106–107, https://doi.org/10.1126/science.aat1884.
4. "How Is Next-Generation Protein Sequencing Democratizing Protein Sequencing?," interview with Jeff Hawkins, CEO of Quantum-Si, News-Medical.net sponsored content, Sept. 5, 2023, https://www.news-medical.net/news/20230905/How-is-next-generation-protein-sequencing-democratizing-protein-sequencing.aspx.
5. Leah V. Schaffer et al., "Identification and Quantification of Proteoforms by Mass Spectrometry" *Proteomics* 19, no. 10 (May 2019): e1800361,

https://doi.org/10.1002/pmic.201800361.

6. Brian D. Reed et al., "Real-Time Dynamic Single-Molecule Protein Sequencing on an Integrated Semiconductor Device," *Science* 378, no. 6616 (Oct. 13, 2022): 186–92, https://doi.org/10.1126/science.abo7651.
7. *Comparison of Protein Sequencing Analysis of CDNF, IL6, and FGF2 on Platinum™ and Mass Spectrometry*, Quantum-Si, 2023, https://www.quantum-si.com/resources/comparison-of-protein-sequencing-analysis-of-cdnf-on-platinum-and-mass-spectrometry/.
8. *Sequencing and Identification of Proteins from Five-Protein Mixtures on Platinum,™* Quantum-Si, 2023, https://www.quantum-si.com/resources/sequencing-and-identification-of-proteins-from-five-protein-mixtures-on-platinum/.
9. *Quantum-Si's Next Generation Protein Sequencing Technology Enables Rapid and Accurate Distinction of Variants of the SARS-CoV-2 Virus*, Quantum-Si, 2023, https://www.quantum-si.com/resources/quantum-sis-next-generation-protein-sequencing-technology-enables-rapid-and-accurate-distinction-of-variants-of-the-sars-cov-2-virus/.
10. Lloyd M. Smith et al., "The Human Proteoform Project: Defining the human proteome," Sci. Adv. 7, no. 46 (Nov. 12, 2021): eabk0734, https://doi.org/10.1126/sciadv.abk0734.
11. *Detecting Arginine Post-Translational Modifications in P38MAPKα and Vimentin Proteins Using Quantum-Si's Next-Generation Protein Sequencing Technology*, Quantum-Si, 2023, https://www.quantum-si.com/resources/detecting-arginine-post-translational-modifications-using-quantum-si-next-generation-protein-sequencing-technology/.

AMERICAN CHEMICAL SOCIETY
1155 Sixteenth Street, NW,
Washington, DC 20036

c&en